Empir Econ (2013) 44:81–109
DOI 10.1007/s00181-010-0354-y

Nonparametric analysis of treatment effects in ordered response models

Stefan Boes

Received: 18 January 2008 / Accepted: 9 October 2009 / Published online: 7 March 2010

Abstract Treatment analyses based on average outcomes do not immediately generalize to the case of ordered responses because the expectation of an ordinally measured variable does not exist. The proposed remedy in this paper is a shift in focus to distributional effects. Assuming a threshold crossing model on both the ordered potential outcomes and the binary treatment variable, and leaving the distribution of error terms and functional forms unspecified, the paper discusses how the treatment effects can be bounded. The construction of bounds is illustrated in a simulated data example.

Keywords Nonparametric bounds · Causal effects · Instrumental variables · Endogenous binary regressor · Partial identification

JEL Classification C14 · C25 · C35

1 Introduction

Suppose one is interested in the effect of a binary treatment D on an ordered response Y. The treatment variable is such that $D = 1$ whenever the treatment is received, and $D = 0$ otherwise. Assume that irrespective of being treated or not the individual faces the same set of mutually exclusive and exhaustive ordered categories of the response variable. Without loss of generality, let $\mathcal{Y} = \{1, 2, \ldots, J\}$ denote this set, where "1" is the smallest outcome and "J" the largest. The assigned values in $\mathcal{Y}$ are entirely

S. Boes
Socioeconomic Institute, University of Zurich, Zuerichbergstrasse 14, 8032 Zurich, Switzerland
e-mail: boes@sts.uzh.ch

S. Boes (✉)
Institute for Quantitative Social Science, Harvard University, 1730 Cambridge Street, Cambridge, MA 02138, USA
e-mail: sboes@iq.harvard.edu

meaningless, as long as they keep the ordering, and are just for notational convenience. It is often useful to think of D as a dummy endogenous variable in the model for Y, provided that the treatment status is determined by self-selected individuals rather than randomly assigned treatment groups.

A number of applications fit into this framework. For example, in medical research the effectiveness of a new drug may be evaluated regarding the patient's health status, the latter measured as *excellent*, *very good*, *good*, *fair*, or *poor*. In educational economics, one may be interested in the effect of out-of-school training programs on student achievements, and the researcher observes final grades *A, B, C, D*. In labor economics, the sorting of workers into public and private sector jobs may be analyzed with respect to their economic performance, measured as *promotion*, *lateral move*, or *demotion*, or one may be interested in the effect of union membership on job satisfaction, the latter recorded from "1" (not at all satisfied) to "10" (completely satisfied). Greene and Hensher (2010) review alternative applications.

The ordinal nature of Y needs to be taken into account when defining treatment effects. Let Y_1 denote the potential outcome with treatment, and let Y_0 denote the potential outcome without it (Neyman 1923; Rubin 1974). With quantitative and binary outcomes, the individual treatment effect $Y_1 - Y_0$ has potential interest. For example, if Y measures wages and D is participation in job training, then $Y_1 - Y_0$ gives the wage difference with and without the training program. With binary outcomes, like participation in the labor force (yes/no), the individual level effect shows the direction of movements caused by the treatment (if any). For ordinal variables, however, such an interpretation does not exist because the distance between outcomes is not defined.

In practice, only one of two potential outcomes can be observed because each individual either receives the treatment, or does not, i.e., $Y = DY_1 + (1 - D)Y_0$. Thus, it is impossible to recover the individual treatment effect, and the literature typically focuses on averages, like the average treatment effect $E(Y_1 - Y_0)$, or the average treatment effect on the treated $E(Y_1 - Y_0|D = 1)$. Under certain assumptions, these parameters can at least partly be recovered from observed data (Heckman and Robb 1985, 1986; Manski 1990, 1994, 1995; Imbens and Angrist 1994; Angrist et al. 1996; Heckman et al. 1999; among many others). With ordinal data, again, the case is different: Any rank preserving recoding of the elements in $\mathcal{Y}$ should not affect the parameters of interest. $E(Y_1)$ and $E(Y_0)$, however, will be affected by such a value conversion so that the concept of averages needs to be replaced by a concept that is insensitive to the definition of $\mathcal{Y}$.

The first contribution of this paper is to present and discuss several treatment parameters that do not rely on the scaling of Y. As for all discrete response variables, each ordinal outcome occurs with a positive probability, and a natural way of analyzing treatment effects is thus in terms of probabilities, or cumulative probabilities, rather than expectations. Most of the parameters described below have been introduced in the previous literature, but generally without particular reference to ordinal outcomes. I will clarify the benefits of using them in that context. Some other parameters, to the best of my knowledge, have not been discussed before.

The key roles in the definition of treatment parameters play the distributions $P(Y_1)$ and $P(Y_0)$ of the potential outcomes. In general, these distributions are not immediately identified from the population distribution of (Y, D) because the counterfactual

probabilities $P(Y_0|D = 1)$ and $P(Y_1|D = 0)$ are not revealed from the sampling process alone. The second contribution of the paper is to bound the counterfactual probabilities, imposing a nonparametric threshold crossing model on both the ordered potential outcomes and the treatment selection. This narrows the identification regions as defined, for example, in Manski (2000, 2003), and is even more informative than the bounds derived, for example, in Heckman and Vytlacil (2001a,b) under a monotone treatment selection assumption.

The approach followed here is closely related to the traditional ordered response literature based on latent variables and threshold crossing mechanisms. For example, parametric models like the ordered probit and the ordered logit model have this structure (McKelvey and Zavoina 1975; McCullagh 1980), but also semiparametric approaches like Klein and Sherman (2002), Bellemare et al. (2002), Coppejans (2007), Lewbel (1997), Lewbel (Unpublished working paper, 2003), and Stewart (2004) impose a threshold crossing model to generate ordinality in the response variable. For a theoretical foundation of ordered response models see Cunha et al. (2007). It therefore seems natural to analyze the implications of such a model structure in a nonparametric bounding analysis. The model is nonparametric in the sense that no distributional assumptions, and no functional form assumptions will be imposed other than the threshold mechanism.

Three papers are related to mine. First, Shaikh and Vytlacil (2005) discuss treatment effect bounds with a binary response variable and a binary treatment. They impose nonparametric threshold crossing models on both the treatment selection and the binary potential outcomes, whereas the model here assumes ordinal potential outcomes. As it will be worked out below, this requires a different bounding strategy, and supplemental interpretations can be given in the extended setting. Second, Scharfstein et al. (2004) bound the distribution of ordinal outcomes, but their model setup is different from mine because they consider two outcome variables where the first is always observed and the second (sequentially following the first) is potentially missing. Third, Li and Tobias (2008) describe Bayesian estimation of treatment effects for ordinal outcomes. They impose more structure on the model than it is imposed here and focus on mean treatment effect parameters (and thus require additional implicit assumptions on the type of ordinal response variable).

The remainder of the paper is structured as follows. In the next section, I present several parameters that are suitable for analyzing treatment effects when the outcome variable is ordinally measured, and I briefly discuss the problem of identification. Section 3 outlines the model. Sections 4 and 5 derive the bounds on the counterfactual probabilities and the treatment parameters, taking into account the nonparametric model structure. Section 6 illustrates the bounds in an artificial data experiment, and Sect. 7 concludes.

2 What treatment parameters are of interest?

The definition of treatment parameters fundamentally depends on the units of measurement of the response variable. Continuous variables received the most attention in the previous literature, and the average treatment effect and the treatment effect on

the treated, or local versions thereof, are well-established and analyzed parameters. If outcomes are measured on an ordinal scale, however, then the prevailing concept of taking averages of the individual level treatment effect is misleading because the difference between outcomes is not defined. In this section, I present a number of parameters that do not rely on the scaling of Y.

2.1 The "standard" parameters

The first parameter is the "probability" counterpart of the average treatment effect (ATE), which can be defined as the probability difference of observing a particular outcome with and without the treatment, formally

$$\Delta_y^{\text{ATE}} \equiv P(Y_1 = y) - P(Y_0 = y) \qquad y = 1, \ldots, J \tag{1}$$

Note that there are indeed J effects, one for each outcome. If the treatment affects responses positively adopting the convention that higher outcomes of Y are in some way "better", then one would expect Δ_y^{ATE} negative for low y and positive for high y. In practice, there may not exist such a clear systematic indicating whether the treatment has a positive or a negative effect, but the shift in focus to probability effects allows for a detailed analysis of the effects of the treatment in all parts of the outcome distribution. Note that Δ_y^{ATE} can be written in terms of expectations, too, using suitable indicator functions: $\Delta_y^{\text{ATE}} \equiv E[\mathbf{1}(Y_1 = y)] - E[\mathbf{1}(Y_0 = y)]$, where $\mathbf{1}(A)$ is one when A is true.

Analogously, the average effect for individuals who actually received the treatment can be defined as the treatment on the treated (TT) parameter

$$\Delta_y^{\text{TT}} \equiv P(Y_1 = y|D = 1) - P(Y_0 = y|D = 1) \qquad y = 1, \ldots, J \tag{2}$$

Both treatment parameters are robust against the particular values assigned to outcomes, but rely on the "same scale" assumption. Yet this assumption is not overly restrictive, as otherwise it would be difficult to compare the Y_1 and the Y_0 distribution. In order to obtain relative (instead of absolute) probability effects, one might normalize (1) and (2) by $P(Y_0 = y)$ and $P(Y_0 = y|D = 1)$, respectively, or the distributions of Y_1. The average treatment effect and the average treatment effect on the treated are the treatment parameters that occur most often in the literature, see for example Manski (2007, Chap. 7), but without particular reference to ordinal response variables. Since they do not rely on the definition of elements in $\mathcal{Y}$, however, they are well-suited to analyze the effect of a treatment in this case.

One may also define other treatment parameters that reflect the ordinal nature of the response variable, such as the local average treatment effect (LATE) of Imbens and Angrist (1994), or the marginal treatment effect (MTE) of Björklund and Moffitt (1987). These parameters are defined for different subgroups of the population. For example, the LATE of Imbens and Angrist (1994) is defined as the average treatment effect for the subgroup of compliers, i.e., those individuals who would comply with the exogenous modification of instruments (where instruments only affect the selec-

tion but not the potential outcomes). Let z_1, z_0 denote two evaluation points of an instrument Z with $P(D = 1|Z = z_1) > P(D = 0|Z = z_0)$. The LATE parameter in terms of probabilities is then given by

$$\Delta_y^{\text{LATE}}(z_1, z_0) \equiv \frac{P(Y = y|Z = z_1) - P(Y = y|Z = z_0)}{P(D = 1|Z = z_1) - P(D = 1|Z = z_0)} \tag{3}$$

This ratio gives the change in the probability distribution for those individuals who would not select into treatment if Z was externally set to a value z such that $P(D = 1|Z = z) \leq P(D = 1|Z = z_0)$, and who would select into treatment if Z was externally set to a value z such that $P(D = 1|Z = z) \geq P(D = 1|Z = z_1)$. An important aspect of the LATE is that it is identified from the population distribution of (Y, D, Z) for all combinations z_1, z_0 with $P(D = 1|Z = z_1) > P(D = 1|Z = z_0)$, which is made explicit in the definition by including z_1 and z_0 in the argument.

A marginal version of the LATE has been introduced in Björklund and Moffitt (1987), see also Heckman (1997) and Heckman and Vytlacil (1999, 2005). It can be defined as the limit of (3) for $P(D = 1|Z = z_0) \to P(D = 1|Z = z_1)$. The MTE gives the change in the probability distribution for those individuals that would just be indifferent between being selected into or out of the treatment if Z was externally set to z such that $P(D = 1|Z = z) = P(D = 1|Z = z_1)$. Heckman and Vytlacil (2001) show how the other treatment parameters can be obtained from the MTE. Both Δ_y^{MTE} and Δ_y^{LATE} are identified from the data, under suitable conditions, and hence identification of Δ_y^{ATE} and Δ_y^{TT} in principle is possible. However, this requires observability of a sufficiently large support of $P(D = 1|Z = z)$, which must not necessarily hold in practice. The analysis below is more general by imposing nonparametric identification regions.

2.2 Some alternatives

While the previous treatment parameters were defined for different groups of the population, the ordinal nature of the response variable allows for a more thorough analysis of the effect on the outcome distribution. In particular, analyzing probabilities rather than expectations provides a much richer set of treatment parameters beyond the common "mean" effects. For example, consider the concept of stochastic order (SO) in two random variables (Mann and Whitney 1947). Let

$$\Delta_y^{\text{SO}} \equiv P(Y_1 \leq y) - P(Y_0 \leq y) \tag{4}$$

If $\Delta_y^{\text{SO}} \leq 0$ for all y, then Y_0 is said to be stochastically smaller than Y_1, i.e., Y_0 tends to have higher probability for low y, and smaller probability for high y compared to Y_1. Analogously, if $\Delta_y^{\text{SO}} \geq 0$ for all y, then Y_0 is said to be stochastically larger than Y_1, and if $\Delta_y^{\text{SO}} = 0$ for all y, then Y_0 and Y_1 are said to be stochastically equivalent. One may also analyze the stochastic order of Y_1 and Y_0 in the subgroup of the treated (SOT)

$$\Delta_y^{\text{SOT}} \equiv P(Y_1 \leq y|D=1) - P(Y_0 \leq y|D=1) \tag{5}$$

where, for example, Y_1 is said to be stochastically larger than Y_0, now conditional on $D=1$, if $\Delta_y^{\text{SOT}} \leq 0$ for all y. If neither of the three cases is true for all y, i.e., Y_1 is not stochastically larger or smaller than, nor equivalent to Y_0, then one may at least analyze the degree of stochastic order starting from $y=1$ moving to $y=J$, or the other way round.

Yet another way to look at the effect of treatment on the outcome distribution, related to the concept of stochastic ordering, is in terms of the relative odds, specifically,

$$\Omega_y \equiv \frac{P(Y_1 > y)/P(Y_1 \leq y)}{P(Y_0 > y)/P(Y_0 \leq y)} \qquad y = 1, \ldots, J-1 \tag{6}$$

for the whole population, and

$$\Omega_y^{\text{T}} \equiv \frac{P(Y_1 > y|D=1)/P(Y_1 \leq y|D=1)}{P(Y_0 > y|D=1)/P(Y_0 \leq y|D=1)} \qquad y = 1, \ldots, J-1 \tag{7}$$

for the treated only. These parameters show the factor by which the ratio of the odds of $Y_1 > y$ relative to $Y_1 \leq y$ in the treatment group change compared to the odds $Y_0 > y$ relative to $Y_0 \leq y$ in the treatment group. With a positive treatment effect, i.e., with the probability of higher outcomes increasing with the receipt of treatment, this factor should be larger than one. If, on the other hand, the treatment effect is negative, then the odds ratio is smaller than one, and if the treatment effect is zero, then the odds ratio is one. Note that there exist $J-1$ odds ratios, one for each category, except for the highest.

2.3 Partial identification of treatment parameters

While the LATE and MTE parameters can be identified from the observed data, but are only defined for a particular subgroup of the population, the other treatment parameters presented above are not immediately identified. The lack of point identification is due to the fact that each individual is observed only in one state, either with or without the treatment. The counterfactual state cannot be inferred from the population distribution of (Y, D), or (Y, D, Z). In order to illustrate the problem, consider the distribution of the potential outcome with treatment $P(Y_1 = y)$. By the law of total probability

$$P(Y_1 = y) = P(Y_1 = y|D=1)P(D=1) + P(Y_1 = y|D=0)P(D=0) \tag{8}$$

The sampling process identifies the probability of treatment selection, $P(D=1)$, and the outcome probability with treatment given treatment has been received, $P(Y_1 = y|D=1) = P(Y = y|D=1)$. The sampling process is uninformative, however, regarding the distribution $P(Y_1 = y|D=0)$, which is the outcome probability with treatment, given the treatment has not been received. In the common terminology such a term is referred to as counterfactual probability. $P(Y_0 = y)$ is not identified either, because the sampling process does not reveal $P(Y_0 = y|D=1)$. However,

one may impose bounds on the unidentified probabilities and thus impose bounds on the potential outcome distribution.

As a starting point and without imposing any assumptions on the data-generating process, it must certainly hold that both counterfactuals, $P(Y_1|D=0)$ and $P(Y_0|D=1)$, are bounded by zero and one.[1] This defines identification regions for the potential outcome distributions, and thus the treatment effects, since all valid $P(Y_1|D=0)$ and $P(Y_0|D=1)$ necessarily yield distributions within the stated bounds (see Manski 2000, 2003 for more details). The task of the rest of the paper is to explore, nonparametrically, the assumptions of a threshold crossing model structure in order to tighten the bounds on the counterfactual distributions. I will first present the model and then discuss how the bounds can be constructed.

3 Model and assumptions

The model for the treatment status and the potential outcomes is a version of the model in Shaikh and Vytlacil (2005) generalized to the case of ordinal outcomes and defined as

$$\begin{aligned} D^* &= s(Z) - \nu \qquad & D &= \mathbf{1}(D^* \geq 0) \\ Y_0^* &= r_0(X) + \varepsilon_0 \qquad & Y_0 &= \sum_{y=1}^{J} y\mathbf{1}(\kappa_{0y-1} < Y_0^* \leq \kappa_{0y}) \\ Y_1^* &= r_1(X) + \varepsilon_1 \qquad & Y_1 &= \sum_{y=1}^{J} y\mathbf{1}(\kappa_{1y-1} < Y_1^* \leq \kappa_{1y}) \end{aligned} \tag{9}$$

where (X, Z) is a random vector of observed covariates, ν, ε_0, and ε_1 are unobserved random variables, and $\mathbf{1}(\cdot)$ is the logical indicator function. The model is a latent index model with latent variables D^*, Y_0^*, and Y_1^*, and a threshold crossing mechanism that generates the treatment status D and the potential outcomes Y_0 and Y_1. The model is nonparametric in the sense that the functional forms of $s(Z)$, $r_0(X)$, and $r_1(X)$ are left unspecified and no parametric assumption on the distribution of $(\varepsilon_0, \varepsilon_1, \nu)$ is made. The model presumes that the error terms and the functions of observable factors are additively separable; see Vytlacil (2002, 2006) for a discussion of this property. Finally, the observed outcome Y is generated according to $DY_1 + (1-D)Y_0$, completing the model.

The model allows for much flexibility in the threshold mechanism since no distributional or functional form assumptions are imposed. In particular, the model does not restrict the shape of treatment effects in a way similar to the single crossing property of probability effects in standard parametric ordered probit and logit models (Boes and Winkelmann 2006), nor does it require a specific model for the threshold parameters in order to relax this property.

[1] For the ease of notation, I will drop the y argument if possible, e.g., $P(Y_1)$ is shorthand notation for $P(Y_1 = y)$.

The treatment is assumed to affect the threshold parameters $(\kappa_{0y},\ \kappa_{1y})$ and the linear indices $(r_{0y},\ r_{1y})$. Depending on the underlying economic model, such a distinction is useful since they might represent different features of the model. For example, Cunha et al. (2007) motivate the ordered threshold model in a choice of goods problem where the linear index measures a consumer's marginal valuation of quality, and the threshold parameters have the interpretation of marginal prices per unit quality. Given such a structure, the treatment might affect prices/quality and/or the valuation of quality.

Models like (9) have recently been employed, for example, in DeVaro (2006) to measure the effect of team production on financial performance (recorded as *about average*, *better than average*, or *a lot better than average*), in Luechinger et al. (2010) to account for self-selection into private and public sector jobs and to estimate the related well-being differentials, and in Munkin and Trivedi (2008) to analyze the effects of different types of health insurance plans on the level of hospital utilization. All these papers use parametric assumptions to identify the treatment parameter of interest, while this paper follows a fully nonparametric approach.

The definition of treatment parameters and the identification regions stated in the previous section still hold conditional on the vector of observed covariates X. In this case, the treatment effects are local (conditional on X), and unconditional effects may be obtained as weighted averages. The model does also include a vector Z that affects the treatment selection. Z may contain all elements of X, and additional elements in Z will generally be referred to as instrumental variables. X may or may not contain an element that is not included in Z. If such an element exists, then this information can be gainfully employed in the bounding analysis. Let $\mathcal{X}$ denote the support of the random vector X, and let $\mathcal{Z}$ denote the support of the random vector Z.

The assumptions imposed on the model are as the ones in Shaikh and Vytlacil (2005), mainly that $(\varepsilon_0, \varepsilon_1, \nu)$ is independent of (Z, X) and that ε_1 and ε_0 have equal distributions conditional on ν. Furthermore, the error terms are assumed to follow continuous and well-behaved distributions with compact support, the distribution of (X, Z) is assumed to have compact support, the functions $r_0(\cdot)$, $r_1(\cdot)$, and $s(\cdot)$ to be continuous, with $s(Z)$ being non-degenerate conditional on X. The extension to ordinal response variables requires two additional assumptions in order to obtain a well-defined model:

(A1) The threshold parameters $\kappa_{0j},\ \kappa_{1j},\ j = 0, \ldots, J$ are fixed and fulfill the order condition $-\infty = \kappa_{00} < \kappa_{01} < \ldots < \kappa_{0J} = \infty$, and $-\infty = \kappa_{10} < \kappa_{11} < \ldots < \kappa_{1J} = \infty$.

(A2) For some $x_0 \in \mathcal{X}$ let $r_0(x_0) = 0$, and for some $x_1 \in \mathcal{X}$ let $r_1(x_1) = 0$.

Assumption (A1) in combination with the model equation explicitly accounts for the order information. The threshold parameters are assumed to be constant and unknown, although the extension to known thresholds (interval data) is possible. In the latter case, knowledge of the thresholds in both treatment statuses is required, unless they are independent of treatment and thus equal. Knowledge of κ_0 and κ_1 will considerably simplify the analysis, and remarks will be given at the appropriate places when the additional information can be used. An immediate extension of the model analyzed here would make the threshold parameters dependent on covariates as well.

Assumption (A2) is an identifying assumption that simplifies exposition and is standard in parametric models. If (A2) is not met, then parametric ordinal response models may only identify location-normalized instead of absolute threshold parameters, i.e., κ_0, κ_1 will be replaced by $\kappa_0 - r_0$ and $\kappa_1 - r_1$, respectively, where r_0, r_1 denote the constants in $r_0(X), r_1(X)$. As it is irrelevant for the following analysis if all thresholds are shifted equally to the right or to the left, (A2) is purely simplifying and does not restrict the analysis in any way.

4 Bounds on treatment effects without covariates

For the ease of exposition, I will first consider bounds on the treatment parameters when no X covariates are available. In this case, the latent potential outcome equations of the model simplify to $Y_1^* = \varepsilon_1$ and $Y_0^* = \varepsilon_0$. The extension to the case when X covariates are present will be separately discussed in Sect. 5. All the proofs are provided in the electronic supplement.

4.1 Bounds under independence

A simple way to construct bounds on the treatment parameters is due to Manski (1990, 1994). Assume that potential outcomes (Y_0, Y_1) are independent of Z, but that treatment selection D varies with Z. For example, if Y measures school grades and D indicates recipiency of a school voucher (intended to support children in low-income families), then Z could be an indicator of winning a lottery for randomly selected families making them eligible for recipiency.

One may interpret such a condition as an exclusion restriction, and Z is an instrumental variable. It is easy to verify that the model assumptions in Sect. 3 imply this condition, but not vice versa. Given independence it must hold that $P(Y_1|Z = z) = P(Y_1)$ for all $z \in \mathcal{Z}$.[2] Furthermore, the smallest of $P(D = 1, Y|Z = z)+P(D = 0|Z = z)$—which is an upper bound of $P(Y_1|Z = z)$—over all $z \in \mathcal{Z}$ may be used as an upper bound for $P(Y_1)$, and the largest of $P(D = 1, Y|Z = z)$— which is a lower bound of $P(Y_1|Z = z)$—over all $z \in \mathcal{Z}$ may be used as a lower bound for $P(Y_1)$. These bounds are commonly referred to as Manski bounds. Analogous arguments hold in order to construct upper and lower bounds for $P(Y_0)$, which then can be used to bound the treatment parameters.

For the treatment on the treated effect note that, in general, $P(Y_0|D = 1, Z) \neq P(Y_0|D = 1)$, i.e., $Y_0|D = 1$ is not independent of Z, as the instrument does affect the treatment status. One option to proceed would be to re-define the treatment on the treated parameter conditional on Z, or conditional on $P(D = 1|Z)$, and then obtain the unconditional parameter by integration. Alternatively, one may rewrite the counterfactual $P(Y_0|D = 1)$ as

[2] In order to save some notation, I will drop the particular value z (or later on x) that is conditioned on if not critical to the given context. It will be implicitly assumed that all expressions are only evaluated over the appropriate support, i.e., at all evaluation points the conditional probabilities exist and are well-defined.

$$P(Y_0 = y|D = 1) = P(D = 1, Y_0 = y)/P(D = 1)$$
$$= [P(Y_0 = y) - P(D = 0, Y_0 = y)]/P(D = 1)$$

by Bayes' theorem and the law of total probability. One may then construct upper and lower bounds on $P(Y_0)$ in the same manner as above. For more details on the construction of the bounds on the average treatment effect and the average treatment effect on the treated I refer to Manski (2003, 2007). Note that the Manski bounds do not exploit the ordinal nature of the response variable, nor do they exploit the threshold crossing structure of the model. The analysis may therefore be applied to any nominal response Y and binary treatment D. The question to be investigated in the following is how such additional assumptions on the structure of the data can be used to improve upon these bounds.

4.2 Bounds under the threshold crossing model structure

The bounding strategy of this section generalizes Heckman and Vytlacil (2001a,b) and Shaikh and Vytlacil (2005) to the case of ordinal potential outcomes. Given the threshold crossing structure of the treatment selection equation and the independence assumption, it follows that for any two evaluation points $z_1, z_0 \in \mathcal{Z}$

$$P(D = 1|Z = z_1) > P(D = 1|Z = z_0) \Leftrightarrow P(s(z_1) \geq \nu) > P(s(z_0) \geq \nu)$$
$$\Leftrightarrow s(z_1) > s(z_0)$$

Furthermore, let

$$z^u = \arg\sup_{z \in \mathcal{Z}} P(D = 1|Z = z)$$
$$z^l = \arg\inf_{z \in \mathcal{Z}} P(D = 1|Z = z)$$

where $\sup\{\cdot\}$ denotes the supremum and $\inf\{\cdot\}$ the infimum of the argument in curly brackets over the values indicated in the subscript. By definition of z^u and z^l it must hold that $s(z^u) \geq s(z)$ and $s(z^l) \leq s(z)$ for all $z \in \mathcal{Z}$. An implication of the threshold crossing treatment selection model (see the Supplementary material for further details) is that the following bounds can be imposed on the average treatment effect and the treatment effect on the treated

$$\Delta_y^{\text{ATE}} \in \left[LBI_y^{\text{ATE}}, UBI_y^{\text{ATE}}\right] \tag{10}$$

with

$$LBI_y^{\text{ATE}} = P(D = 1, Y = y|Z = z^u) - P(D = 1|Z = z^l)$$
$$-P(D = 0, Y = y|Z = z^l)$$

$$UBI_y^{\text{ATE}} = P(D=1, Y=y|Z=z^u) + P(D=0|Z=z^u) - P(D=0, Y=y|Z=z^l)$$

and

$$\Delta_y^{\text{TT}} \in \left[LBI_y^{\text{TT}}, UBI_y^{\text{TT}}\right] \tag{11}$$

with

$$\begin{aligned} LBI_y^{\text{TT}} &= [P(Y=y) - P(D=1|Z=z^l) \\ &\quad - P(D=0, Y=y|Z=z^l)]/P(D=1) \\ UBI_y^{\text{TT}} &= [P(Y=y) - P(D=0, Y=y|Z=z^l)]/P(D=1) \end{aligned}$$

The bounds are sharp, given the assumptions (see Heckman and Vytlacil (2001a) for details about the proof). The bounds can be readily evaluated once z^u and z^l are determined. It is also possible to calculate their width. For the average effect the width is given by $P(D=0|Z=z^u) + P(D=1|Z=z^l)$, which is the smaller the larger the variation in D over the support of Z. For the treatment on the treated parameter the width is given by $P(D=1|Z=z^l)/P(D=1)$ which is the smaller the larger the overall fraction of the selected, i.e., the larger $P(D=1)$, and the smaller $P(D=1|Z=z^l)$. Both are smaller than one given that treatment selection varies with Z, i.e., the independence assumption together with the threshold crossing treatment selection is informative.

Now consider the combination of threshold crossing treatment selection and the threshold model for the potential outcomes. Let

$$\text{sgn}(a) = \begin{cases} -1 & \text{if} \quad a < 0 \\ 0 & \text{if} \quad a = 0 \\ 1 & \text{if} \quad a > 0 \end{cases}$$

denote the sign function, and consider the following lemma:

Lemma 1 *Assume that* (Y_0, Y_1, D) *are generated according to model (9), and assume that all conditions outlined in Sect. 3 are fulfilled. Then for any two evaluation points* z_1, z_0 *with* $P(D=1|Z=z_1) > P(D=1|Z=z_0)$,

$$sgn[P(Y \leq y|Z=z_1) - P(Y \leq y|Z=z_0)] = sgn(\kappa_{1y} - \kappa_{0y}) \equiv \delta_y$$

so that δ_y *can take three values* $-1, 0, 1$ *depending on whether the difference* $\kappa_{1y} - \kappa_{0y}$ *is negative, zero, or positive, respectively.*

The intuition for Lemma 1 is as follows. Since Z shifts the probability of treatment selection (by assumption), Z will generally also cause a shift in the cumulative probabilities of the observed Y (indirectly through D, unless the outcome processes in both states are the same). Due to the threshold crossing model structure, the direction

of the latter shift can be directly related to the relative magnitude of the threshold parameters. Lemma 1 is analogous to Lemma 4.2 of Shaikh and Vytlacil (2005), but now with respect to the properties of ordinal potential outcomes.

Although it is not possible, without further assumptions, to identify the absolute magnitude of threshold parameters, information on the relative magnitude can already be used to tighten the bounds on the unidentified probabilities $P(Y_0|D = 1, Z)$ and $P(Y_1|D = 0, Z)$. Rewrite $P(Y_1|D = 0, Z)$ as

$$P(Y_1 = y|D = 0, Z) = P(Y_1 \leq y|D = 0, Z) - P(Y_1 \leq y - 1|D = 0, Z)$$

which follows from the ordinal nature of Y. Furthermore, the difference

$$\begin{aligned} &P(Y_1 \leq y|D = 0, Z) - P(Y_0 \leq y|D = 0, Z) \\ &\quad = P(\varepsilon_1 \leq \kappa_{1y}|\nu > s(z)) - P(\varepsilon_0 \leq \kappa_{0y}|\nu > s(z)) \\ &\quad = P(\varepsilon \leq \kappa_{1y}|\nu > s(z)) - P(\varepsilon \leq \kappa_{0y}|\nu > s(z)) \end{aligned} \tag{12}$$

has the same sign as $\kappa_{1y} - \kappa_{0y}$, and $\delta_y \equiv \text{sgn}(\kappa_{1y} - \kappa_{0y})$ is identified by Lemma 1. This must hold for all possible outcomes y, so that by the model assumptions, the sign of the difference

$$\begin{aligned} &P(Y_1 \leq y - 1|D = 0, Z) - P(Y_0 \leq y - 1|D = 0, Z) \\ &\quad = P(\varepsilon_1 \leq \kappa_{1y-1}|\nu > s(z)) - P(\varepsilon_0 \leq \kappa_{0y-1}|\nu > s(z)) \\ &\quad = P(\varepsilon \leq \kappa_{1y-1}|\nu > s(z)) - P(\varepsilon \leq \kappa_{0y-1}|\nu > s(z)) \end{aligned} \tag{13}$$

equals $\delta_{y-1} \equiv \text{sgn}(\kappa_{1y-1} - \kappa_{0y-1})$.

The strategy to bound the unidentified probabilities is a pairwise comparison of terms in

$$\begin{aligned} &P(Y_1 = y|D = 0, Z) - P(Y_0 = y|D = 0, Z) \\ &\quad = [P(Y_1 \leq y|D = 0, Z) - P(Y_1 \leq y - 1|D = 0, Z)] \\ &\qquad -[P(Y_0 \leq y|D = 0, Z) - P(Y_0 \leq y - 1|D = 0, Z)] \\ &\quad = [P(Y_1 \leq y|D = 0, Z) - P(Y_0 \leq y|D = 0, Z)] \\ &\qquad -[P(Y_1 \leq y - 1|D = 0, Z) - P(Y_0 \leq y - 1|D = 0, Z)] \end{aligned} \tag{14}$$

With three different outcomes of δ_y and δ_{y-1}, there are nine possibilities to consider in total. The following lemma states and summarizes the results for both unidentified probabilities:

Lemma 2 *Assume that* (Y_0, Y_1, D) *are generated according to model (9), and assume that all conditions outlined in Sect. 3 are fulfilled. Then,*

$$
\begin{aligned}
&\delta_y > \delta_{y-1} \\
&\quad \Leftrightarrow P(Y_1 = y|D = 0, Z) > P(Y_0 = y|D = 0, Z) = P(Y = y|D = 0, Z) \\
&\qquad P(Y_0 = y|D = 1, Z) < P(Y_1 = y|D = 1, Z) = P(Y = y|D = 1, Z) \\
&\delta_y = \delta_{y-1} = 0 \\
&\quad \Leftrightarrow P(Y_1 = y|D = 0, Z) = P(Y_0 = y|D = 0, Z) = P(Y = y|D = 0, Z) \\
&\qquad P(Y_0 = y|D = 1, Z) = P(Y_1 = y|D = 1, Z) = P(Y = y|D = 1, Z) \\
&\delta_y < \delta_{y-1} \\
&\quad \Leftrightarrow P(Y_1 = y|D = 0, Z) < P(Y_0 = y|D = 0, Z) = P(Y = y|D = 0, Z) \\
&\qquad P(Y_0 = y|D = 1, Z) > P(Y_1 = y|D = 1, Z) = P(Y = y|D = 1, Z)
\end{aligned}
$$

If $\delta_y = \delta_{y-1} = \pm 1$*, then the signs of the differences* $P(Y_1 = y|D = 0, Z) - P(Y_0 = y|D = 0, Z)$ *and* $P(Y_0 = y|D = 1, Z) - P(Y_1 = y|D = 1, Z)$ *are indeterminate.*

Lemma 2 uses the information revealed by Lemma 1 to impose bounds on the counterfactual probabilities tighter than the logical unit range. In particular, knowledge of the relative magnitude of the threshold parameters in both outcome processes allows to evaluate the relative magnitude of the two potential outcome distributions conditional on the selection status, one of which can be identified from the observed data and thus serves as a natural bound. The bounds in Lemma 2 correspond to those in Manski and Pepper (2000) obtained under a monotone instrumental variables and a monotone treatment response assumption.

Without loss of generality, take the two evaluation points z^l and z^u with $s(z^u) > s(z^l)$, and apply Lemma 1 to identify the relative magnitude of threshold parameters. Suppose, the information is revealed that $\delta_y > \delta_{y-1}$. Then $P(Y = y|D = 0, Z)$ can be used as a lower bound for $P(Y_1 = y|D = 0, Z)$ instead of zero, and $P(Y = y|D = 1, Z)$ can be used as an upper bound for $P(Y_0 = y|D = 1, Z)$ instead of one. Bounds on $P(Y_1|Z)$ and $P(Y_0|Z)$ are thus given by

$$
\begin{aligned}
P(Y = y|Z) \leq P(Y_1 = y|Z) &\leq P(D = 1, Y = y|Z) + P(D = 0|Z) \\
P(D = 0, Y = y|Z) \leq P(Y_0 = y|Z) &\leq P(Y = y|Z)
\end{aligned}
$$

If alternatively the information is revealed that $\delta_y < \delta_{y-1}$, then the bounds on $P(Y_1|Z)$, $P(Y_0|Z)$ can be derived as

$$
\begin{aligned}
P(D = 1, Y = y|Z) \leq P(Y_1 = y|Z) &\leq P(Y = y|Z) \\
P(Y = y|Z) \leq P(Y_0 = y|Z) &\leq P(D = 1|Z) + P(D = 1, Y = y|Z)
\end{aligned}
$$

If upper and lower treated and non-treated thresholds are equal, then the outcome of Y does not vary with the treatment status because the cumulative probabilities are unchanged, and the unidentified probabilities become identified, i.e., $P(Y_1|Z) = P(Y|Z) = P(Y_0|Z)$. The bounds imposed by Lemma 2 thus depend on the category

under consideration, i.e., one may have $\delta_y > \delta_{y-1}$, but $\delta_{y+1} < \delta_y$, such that the restrictions on counterfactual probabilities in category y are different from the restrictions in category $y+1$. If Lemmas 1 and 2 do not reveal further information on the counterfactual probabilities, then the lower bound zero and the upper bound one on $P(Y_1|D=0,Z)$ and $P(Y_0|D=1,Z)$ still apply.

The model assumptions imply that $P(Y_1|Z) = P(Y_1)$ and $P(Y_0|Z) = P(Y_0)$. $P(Y_1)$ and $P(Y_0)$ must therefore necessarily lie within the intersection over all possible z so that lower bounds can be replaced by supremum expressions, and upper bounds can be replaced by infimum expressions. The Supplementary material shows how these expressions can be simplified. The following proposition uses the bounds on $P(Y_0)$ and $P(Y_1)$ under the threshold crossing model structure of treatment selection and potential outcomes to bound the average treatment and the treatment on the treated parameters:

Proposition 1 *Assume that (Y_0, Y_1, D) are generated according to model (9), and assume that all conditions outlined in Sect. 3 are fulfilled. Then,*

$$\Delta_y^{\text{ATE}} \in \left[LB2_y^{\text{ATE}}, UB2_y^{\text{ATE}} \right] \tag{15}$$

with

$$LB2_y^{\text{ATE}} = \begin{cases} P(Y=y|Z=z^u) - P(Y=y|Z=z^l) & \text{if } \delta_y > \delta_{y-1} \\ 0 & \text{if } \delta_y = \delta_{y-1} = 0 \\ LB1_y^{\text{ATE}} & \text{if } \delta_y < \delta_{y-1} \\ LB1_y^{\text{ATE}} & \text{if } \delta_y = \delta_{y-1} = \pm 1 \end{cases}$$

$$UB2_y^{\text{ATE}} = \begin{cases} UB1_y^{\text{ATE}} & \text{if } \delta_y > \delta_{y-1} \\ 0 & \text{if } \delta_y = \delta_{y-1} = 0 \\ P(Y=y|Z=z^u) - P(Y=y|Z=z^l) & \text{if } \delta_y < \delta_{y-1} \\ UB1_y^{\text{ATE}} & \text{if } \delta_y = \delta_{y-1} = \pm 1 \end{cases}$$

and

$$\Delta_y^{\text{TT}} \in \left[LB2_y^{\text{TT}}, UB2_y^{\text{TT}} \right] \tag{16}$$

with

$$LB2_y^{\text{TT}} = \begin{cases} [P(Y=y) - P(Y=y|Z=z^l)]/P(D=1) & \text{if } \delta_y > \delta_{y-1} \\ 0 & \text{if } \delta_y = \delta_{y-1} = 0 \\ LB1_y^{\text{TT}} & \text{if } \delta_y < \delta_{y-1} \\ LB1_y^{\text{TT}} & \text{if } \delta_y = \delta_{y-1} = \pm 1 \end{cases}$$

$$UB2_y^{\text{TT}} = \begin{cases} UB1_y^{\text{TT}} & \text{if } \delta_y > \delta_{y-1} \\ 0 & \text{if } \delta_y = \delta_{y-1} = 0 \\ \left[P(Y=y) - P(Y=y|Z=z^l)\right]/P(D=1) & \text{if } \delta_y < \delta_{y-1} \\ UB1_y^{\text{TT}} & \text{if } \delta_y = \delta_{y-1} = \pm 1 \end{cases}$$

The bounds are sharp given the assumptions. For known threshold parameters (interval data), (15) and (16) still hold, but δ_y and δ_{y-1} can a-priori be determined and there is no uncertainty about the four cases.

Note that the width of the bounds in (15) and (16) is at maximum the same and in many cases smaller than the width of the bounds in (10) and (11). If $\delta_y > \delta_{y-1}$, then the upper bound in (15) corresponds to the upper bound in (10), but the lower bound in (15) is larger than the lower bound in (10), since $L\mathcal{B}2_y^{\text{ATE}} - L\mathcal{B}1_y^{\text{ATE}}$ equals

$$P(D = 0, Y = y|Z = z^u) - P(D = 1, Y = y|Z = z^l) + P(D = 1|Z = z^l) > 0$$

With the same argument, if $\delta_y < \delta_{y-1}$, then the lower bounds in (15) and (10) are the same, but the upper bound in (15) is lower than that in (10).

Analogously, for the treatment on the treated parameter and a positive sign of the difference $\delta_y - \delta_{y-1}$, the lower bound in (16) is larger than the lower bound in (11), i.e., $L\mathcal{B}2_y^{\text{TT}} - L\mathcal{B}1_y^{\text{TT}} > 0$, with the upper bounds unchanged, and if $\delta_y - \delta_{y-1}$ is negative, then the upper bound in (16) is lower than the upper bound in (11), with the lower bounds unchanged. If $\delta_y = \delta_{y-1} = 0$, then both treatment parameters become point-identified to be zero. Only in the case $\delta_y = \delta_{y-1} = \pm 1$, the width of the bounds does not change and the threshold mechanism is uninformative on the treatment parameters.

Note that unlike for the bounds constructed before, the sign of Δ_y^{ATE} and Δ_y^{TT} as bounded by Proposition 1 can be identified if $\delta_y \lessgtr \delta_{y-1}$ or $\delta_y = \delta_{y-1} = 0$. This follows because the lower bounds $L\mathcal{B}2_y^{\text{ATE}}$ and $L\mathcal{B}2_y^{\text{TT}}$ of both treatment parameters are positive in the case $\delta_y > \delta_{y-1}$, and in the case $\delta_y < \delta_{y-1}$ the upper bounds $U\mathcal{B}2_y^{\text{ATE}}$ and $U\mathcal{B}2_y^{\text{TT}}$ are negative. Finally, if $\delta_y = \delta_{y-1} = 0$, then the treatment effects is point-identified to be zero.

The final remark on (15) and (16) is related to the case of known thresholds. Given the assumptions of the model and provided that no X covariates are available, the only way that treated and non-treated individuals may differ are the threshold parameters. If the thresholds do not vary by the treatment status, and are thus equal, then $\delta_y = \delta_{y-1} = 0$ in all cases and the treatment parameters are point-identified to be zero, as predicted by Proposition 1.

5 Bounds in the presence of covariates

I now turn to the case when X covariates are available and to the full model (9). By the preceding discussion, it is straightforward to show that $P(Y_1|X)$ and $P(Y_0|X)$ are only partially identified, and so are the treatment parameters. The offending terms are, as before, the probabilities $P(D = 0, Y_1|X)$ and $P(D = 1, Y_0|X)$, respectively. All the results derived before are trivially extended to X conditioned on. There is, however, a potential source of narrowing the bounds, given that X varies conditional on Z, i.e., there exists at least one element in X that is not included in Z. This extra-variation can be explored as follows. Consider the following modified version of Lemma 1:

Lemma 3 *Assume that (Y_0, Y_1, D) are generated according to model (9), and assume that all conditions outlined in Sect. 3 are fulfilled. Then for any evaluation points*

x_0, x_1, z_0, z_1 *with* $P(D = 1|X = x_j, Z = z_1) > P(D = 1|X = x_j, Z = z_0)$, $j = 0, 1$,

$$\begin{aligned} &sgn\{[P(D = 1, Y \leq y|X = x_1, Z = z_1) \\ &\quad - P(D = 1, Y \leq y|X = x_1, Z = z_0)] \\ &\quad - [P(D = 0, Y \leq y|X = x_0, Z = z_0) \\ &\quad - P(D = 0, Y \leq y|X = x_0, Z = z_1)]\} \\ &= sgn(\kappa_{1y}(x_1) - \kappa_{0y}(x_0)) \equiv \delta_y(x_1, x_0) \end{aligned}$$

so that $\delta_y(x_1, x_0)$ *can take three values* $-1, 0, 1$ *depending on whether the difference between* $\kappa_{1y}(x_1) \equiv \kappa_{1y} - r_1(x_1)$ *and* $\kappa_{0y}(x_0) \equiv \kappa_{0y} - r_0(x_0)$ *is negative, zero, or positive, respectively.*

Lemma 3 uses the extra-variation in X to explore the relative magnitude of $\kappa_{0y}(x_0)$ (evaluated at x_0) and $\kappa_{1y}(x_1)$ (evaluated at $x_1 \neq x_0$). Without the extra-variation, it would only be possible to identify the relative magnitude of the linear indices (including the thresholds) in both treatment states if evaluated at the same point $x_1 = x_0$. The extra-variation in X and the result in Lemma 3 can be used to obtain bounds on $P(D = 0, Y_1 = y|X, Z)$ and $P(D = 1, Y_0 = y|X, Z)$. Consider the former probability, and recall that

$$P(D = 0, Y_1 = y|X, Z) = P(D = 0, Y_1 \leq y|X, Z) - P(D = 0, Y_1 \leq y - 1|X, Z)$$

by the ordinal nature of Y. Take the former cumulative probability, evaluated at x_1, and subtract the identified probability $P(D = 0, Y \leq y|X, Z)$ evaluated at x_0 to obtain

$$\begin{aligned} &P(D = 0, Y_1 \leq y|X = x_1, Z) - P(D = 0, Y_0 \leq y|X = x_0, Z) \\ &= P(\varepsilon \leq \kappa_{1y}(x_1), \nu > s(z)) - P(\varepsilon \leq \kappa_{0y}(x_0), \nu > s(z)) \end{aligned}$$

The sign of the (unidentified) difference depends on the sign of the difference $\kappa_{1y}(x_1) - \kappa_{0y}(x_0)$, which is identified by Lemma 3. Thus, if $\delta_y(x_1, x_0) > 0$, and hence $\kappa_{1y}(x_1) > \kappa_{0y}(x_0)$, then the above difference will be positive. If $\delta_y(x_1, x_0) < 0$, then the above difference will be negative, and if $\delta_y(x_1, x_0) = 0$, then $P(D = 0, Y_1 \leq y|X, Z) = P(D = 0, Y_0 \leq y|X, Z)$ becomes point-identified. Since Lemma 3 holds for all $y \in \mathcal{Y}$, analogous arguments hold for category $y - 1$. A pairwise comparison of terms in

$$\begin{aligned} &P(D = 0, Y_1 = y|X = x_1, Z) - P(D = 0, Y_0 = y|X = x_0, Z) \\ &\quad = P(D = 0, Y_1 \leq y|X = x_1, Z) - P(D = 0, Y_1 \leq y - 1|X = x_1, Z) \\ &\qquad - [P(D = 0, Y_0 \leq y|X = x_0, Z) - P(D = 0, Y_0 \leq y - 1|X = x_0, Z)] \\ &\quad = P(D = 0, Y_1 \leq y|X = x_1, Z) - P(D = 0, Y_0 \leq y|X = x_0, Z) \\ &\qquad - [P(D = 0, Y_1 \leq y - 1|X = x_1, Z) - P(D = 0, Y_0 \leq y - 1|X = x_0, Z)] \\ &\quad = P(\varepsilon \leq \kappa_{1y}(x_1), \nu > s(z)) - P(\varepsilon \leq \kappa_{0y}(x_0), \nu > s(z)) \\ &\qquad - [P(\varepsilon \leq \kappa_{1y-1}(x_1), \nu > s(z)) - P(\varepsilon \leq \kappa_{0y-1}(x_0), \nu > s(z))] \end{aligned} \tag{17}$$

may thus be used to obtain bounds on the unidentified counterfactual probabilities. For example, if Lemma 3 reveals the information that $\delta_y(x_1, x_0) > \delta_{y-1}(x_1, x_0)$, then the difference between the former two probabilities after the last equality in (17) must be larger than the difference between the latter two, so that the overall sign is positive, and $P(D = 0, Y_0 = y|X = x_0, Z)$ can be used as lower bound for $P(D = 0, Y_1 = y|X = x_1, Z)$ instead of zero. By the same arguments, bounds on the counterfactual probability $P(D = 0, Y_0 = y|X, Z)$ can be obtained. The following lemma summarizes and states the results:

Lemma 4 *Assume that (Y_0, Y_1, D) are generated according to model (9), and assume that all conditions outlined in Sect. 3 are fulfilled. Then,*

$$
\begin{aligned}
(a)\quad \delta_y(x, \tilde{x}) &> \delta_{y-1}(x, \tilde{x}) \\
&\Leftrightarrow P(D = 0, Y_1 = y|X = x, Z) > P(D = 0, Y = y|X = \tilde{x}, Z) \\
\delta_y(x, \tilde{x}) &= \delta_{y-1}(x, \tilde{x}) = 0 \\
&\Leftrightarrow P(D = 0, Y_1 = y|X = x, Z) = P(D = 0, Y = y|X = \tilde{x}, Z) \\
\delta_y(x, \tilde{x}) &< \delta_{y-1}(x, \tilde{x}) \\
&\Leftrightarrow P(D = 0, Y_1 = y|X = x, Z) < P(D = 0, Y = y|X = \tilde{x}, Z)
\end{aligned}
$$

If $\delta_y(x, \tilde{x}) = \delta_{y-1}(x, \tilde{x}) = \pm 1$, then the sign of the difference $P(D = 0, Y_1 = y|X = x, Z) - P(D = 0, Y_0 = y|X = \tilde{x}, Z)$ is indeterminate. And,

$$
\begin{aligned}
(b)\quad \delta_y(\tilde{x}, x) &> \delta_{y-1}(\tilde{x}, x) \\
&\Leftrightarrow P(D = 1, Y_0 = y|X = x, Z) < P(D = 1, Y = y|X = \tilde{x}, Z) \\
\delta_y(\tilde{x}, x) &= \delta_{y-1}(\tilde{x}, x) = 0 \\
&\Leftrightarrow P(D = 1, Y_0 = y|X = x, Z) = P(D = 1, Y = y|X = \tilde{x}, Z) \\
\delta_y(\tilde{x}, x) &< \delta_{y-1}(\tilde{x}, x) \\
&\Leftrightarrow P(D = 1, Y_0 = y|X = x, Z) > P(D = 1, Y = y|X = \tilde{x}, Z)
\end{aligned}
$$

If $\delta_y(\tilde{x}, x) = \delta_{y-1}(\tilde{x}, x) = \pm 1$, then the sign of the difference $P(D = 1, Y_0 = y|X = x, Z) - P(D = 1, Y_1 = y|X = \tilde{x}, Z)$ is indeterminate.

Lemma 4 holds for all evaluation points $\tilde{x}$ in the support of X. As there might be some evaluation points $\tilde{x}$ for that $\delta_y(x, \tilde{x}) > \delta_{y-1}(x, \tilde{x})$, and some other evaluation points $\tilde{x}$ for that $\delta_y(x, \tilde{x}) < \delta_{y-1}(x, \tilde{x})$, or $\delta_y(x, \tilde{x}) = \delta_{y-1}(x, \tilde{x}) = 1$, one can use this information to obtain bounds that are tighter than the bounds obtained without the extra-variation in X. Let

$$
\begin{aligned}
\mathcal{X}_0^l(x_1) &= \{x_0 : \delta_y(x_1, x_0) > \delta_{y-1}(x_1, x_0)\} \\
\mathcal{X}_0^u(x_1) &= \{x_0 : \delta_y(x_1, x_0) < \delta_{y-1}(x_1, x_0)\}
\end{aligned}
$$

and

$$\mathcal{X}_1^l(x_0) = \{x_1 : \delta_y(x_1, x_0) < \delta_{y-1}(x_1, x_0)\}$$
$$\mathcal{X}_1^u(x_0) = \{x_1 : \delta_y(x_1, x_0) > \delta_{y-1}(x_1, x_0)\}$$

It is made explicit in the definition of sets that these are either over x_0 for x_1 fixed (and thus are a function of x_1), or over x_1 for x_0 fixed (and thus are a function of x_0). Bounds on $P(D = 0, Y_1 = y|X, Z)$, conditional on all values z in the support of Z can then be derived as

$$\begin{aligned}&\sup_{\tilde{x}\in\mathcal{X}_0^l(x)} \{P(D = 0, Y = y|X = \tilde{x}, Z)\}\\ &\quad \leq P(D = 0, Y_1 = y|X = x, Z) \leq\\ &\inf_{\tilde{x}\in\mathcal{X}_0^u(x)} \{P(D = 0, Y = y|X = \tilde{x}, Z)\}\end{aligned}$$

If there exists $\tilde{x}$ such that $\delta_y(x, \tilde{x}) = \delta_{y-1}(x, \tilde{x}) = 0$ (for x fixed), then point-identification of the counterfactual probability follows. If no such $\tilde{x}$ exists, and no $\tilde{x}$ for that Lemma 4 yields tighter bounds than the unit range, then $\mathcal{X}_0^l$ and $\mathcal{X}_0^u$ are empty and it is understood that the bounds zero and one still apply. Analogously, for $P(D = 1, Y_0 = y|X, Z)$ the bounds can be derived as

$$\begin{aligned}&\sup_{\tilde{x}\in\mathcal{X}_1^l(x)} \{P(D = 1, Y = y|X = \tilde{x}, Z)\}\\ &\quad \leq P(D = 1, Y_0 = y|X = x, Z) \leq\\ &\inf_{\tilde{x}\in\mathcal{X}_1^u(x)} \{P(D = 1, Y = y|X = \tilde{x}, Z)\}\end{aligned}$$

with point-identification $P(D = 1, Y_0 = y|X = x, Z) = P(D = 1, Y = y|X = \tilde{x}, Z)$ if there exists $\tilde{x}$ such that $\delta_y(\tilde{x}, x) = \delta_{y-1}(\tilde{x}, x) = 0$, and bounds zero and one if $\mathcal{X}_1^l$ and $\mathcal{X}_1^u$ are empty.

Replacing the bounds for the counterfactual probabilities in the expressions for $P(Y_1|X, Z)$ and $P(Y_0|X, Z)$ and following the same arguments as under the independence assumption, yields

$$\begin{aligned}LB_y^1(x) \equiv &\sup_{z\in\mathcal{Z}} \{P(D = 1, Y = y|X = x, Z = z)\\ &+ \sup_{\tilde{x}\in\mathcal{X}_0^l(x)} \{P(D = 0, Y = y|X = \tilde{x}, Z = z)\}\}\\ &\leq P(Y_1 = y|X = x) \leq\\ UB_y^1(x) \equiv &\inf_{z\in\mathcal{Z}} \{P(D = 1, Y = y|X = x, Z = z)\\ &+ \inf_{\tilde{x}\in\mathcal{X}_0^u(x)} \{P(D = 0, Y = y|X = \tilde{x}, Z = z)\}\}\end{aligned} \tag{18}$$

and

$$
\begin{aligned}
L\mathcal{B}_y^0(x) \equiv \sup_{z\in\mathcal{Z}} \Bigg\{ & \sup_{\tilde{x}\in\mathcal{X}_1^l(x)} \{P(D=1, Y=y|X=\tilde{x}, Z=z)\} \\
& + P(D=0, Y=y|X=x, Z=z) \Bigg\} \\
\leq P(Y_0 = y|X=x) \leq \\
U\mathcal{B}_y^0(x) \equiv \inf_{z\in\mathcal{Z}} \Bigg\{ & \inf_{\tilde{x}\in\mathcal{X}_1^u(x)} \{P(D=1, Y=y|X=\tilde{x}, Z=z)\} \\
& + P(D=0, Y=y|X=x, Z=z) \Bigg\} \qquad (19)
\end{aligned}
$$

The following proposition uses the bounds in (18) and (19) under the threshold crossing model structure and the full model to impose bounds on the average treatment effect and the average treatment effect on the treated conditional on X:

Proposition 2 *Assume that (Y_0, Y_1, D) are generated according to model (9), and assume that all conditions outlined in Sect. 3 are fulfilled. Then,*

$$
\Delta_y^{\text{ATE}}(x) \in \left[L\mathcal{B}2_y^{\text{ATE}}(x), U\mathcal{B}2_y^{\text{ATE}}(x) \right] \qquad (20)
$$

with

$$
\begin{aligned}
L\mathcal{B}2_y^{\text{ATE}}(x) &= L\mathcal{B}_y^1(x) - U\mathcal{B}_y^0(x) \\
U\mathcal{B}2_y^{\text{ATE}}(x) &= U\mathcal{B}_y^1(x) - L\mathcal{B}_y^0(x)
\end{aligned}
$$

and

$$
\Delta_y^{\text{TT}}(x) \in \left[L\mathcal{B}2_y^{\text{TT}}(x), U\mathcal{B}2_y^{\text{TT}}(x) \right] \qquad (21)
$$

with

$$
\begin{aligned}
L\mathcal{B}2_y^{\text{TT}}(x) &= [P(Y=y|X=x) - U\mathcal{B}_y^0(x)]/P(D=1|X=x) \\
U\mathcal{B}2_y^{\text{TT}}(x) &= [P(Y=y|X=x) - L\mathcal{B}_y^0(x)]/P(D=1|X=x)
\end{aligned}
$$

For $(X, Z) \in \mathcal{X} \times \mathcal{Z}$ the bounds are sharp given the assumptions.

The bounds imposed by Proposition 2 depend on the amount of variation in X conditional on Z, and therefore it is difficult to make a general statement about their properties. However, two important conclusions can be drawn. First, if X does not vary conditional on Z, then the bounds in (20) and (21) simplify to the bounds in (15) and (16) with X conditioned on, but there is no possibility to further narrow the bounds. The reason is that if X is degenerate conditional on Z, then there exists only one $\tilde{x} = x$ in Lemma 4, which then becomes equivalent to Lemma 2 conditional on X. Thus, the cases $\delta_y(x, x) \lesseqgtr \delta_{y-1}(x, x)$, $\delta_y(x, x) = \delta_{y-1}(x, x) = 0$ allow to impose new upper

and/or lower bounds on the counterfactual probabilities, if $\delta_y(x,x) = \delta_{y-1}(x,x) \pm 1$, then the bounds zero and one still apply, and as a consequence, the bounds in Proposition 2 collapse to those in Proposition 1 (conditional on X).

Second, the sign of the treatment effects may be identified using the bounds in Proposition 2. First consider the bounds in (20) and suppose that $\delta_y(x,x) > \delta_{y-1}(x,x)$, $x \in \mathcal{X}_0^l(x)$ and $x \in \mathcal{X}_1^u(x)$ by Lemma 3. For the lower bound it must hold that

$$\begin{aligned} &LB_y^1(x) - UB_y^0(x) \\ &\quad = \sup_{z\in Z}\{P(D=1,Y=y|X=x,Z=z) + P(D=0,Y=y|X=x,Z=z)\} \\ &\qquad - \inf_{z\in Z}\{P(D=1,Y=y|X=x,Z=z) + P(D=0,Y=y|X=x,Z=z)\} \\ &\quad = P(Y=y|X=x,Z=z^u) - P(Y=y|X=x,Z=z^l) > 0 \end{aligned}$$

which follows by Lemma 4 and the definition of z^u and z^l. The inequality holds for $\tilde{x} = x$, if other $\tilde{x} \in \mathcal{X}_0^l(x)$ and $\tilde{x} \in \mathcal{X}_1^u(x)$ exist, then $LB_y^1(x)$ may get larger but never can get smaller by the supremum condition, and $UB_y^0(x)$ may get smaller but never can get larger by the infimum condition, so that the inequality will still hold, and the lower bound in (20) will strictly be positive. By similar arguments, one can show that the upper bound $UB_y^1(x) - LB_y^0(x)$ is negative for $\tilde{x} = x$, and will always be negative for all $\tilde{x} \in \mathcal{X}_0^u(x)$ and $\tilde{x} \in \mathcal{X}_1^l(x)$ other than x if $\delta_y(x,x) < \delta_{y-1}(x,x)$. In the case that $\delta_y(x,x) = \delta_{y-1}(x,x) = 0$, then the counterfactual probabilities become identified by Lemma 4, and the average treatment effect is point-identified to be zero. Analogous arguments hold for the treatment on the treated parameter.

6 An artificial data example

The construction and some of the properties of the bounds are illustrated using a simulated data example. The model is specified as follows:

$$\begin{aligned} D &= \mathbf{1}(\alpha Z + \nu \geq 0) \\ Y_0 &= \sum_{y=1}^{3} y\mathbf{1}(\kappa_{0y-1} < \varepsilon \leq \kappa_{0y}), \quad Y_1 = \sum_{y=1}^{3} y\mathbf{1}(\kappa_{1y-1} < \gamma X + \varepsilon \leq \kappa_{1y}) \end{aligned}$$

where $\kappa_{11} = \kappa_{01} = 0$ and $\kappa_{12} = \kappa_{02} + \beta = 1 + \beta$. (ε, ν) are bivariate standard normally distributed with correlation 0.2. (X, Z) are both binary, taking value one if an underlying latent variable is positive, with the two latent variables following a bivariate standard normal distribution with correlation 0.5.

I confine myself to the construction of bounds on the average treatment effect and consider three different scenarios. Scenario 1 assumes that $\gamma = 0$ and β is varied from -0.5 to 0.5. Thus, there are no X covariates and the treatment does only affect the thresholds. The effect is asymmetric in the sense that only the upper threshold is affected. Scenario 2 assumes that $\beta = 0$ and γ is varied from -0.5 to 0.5. The

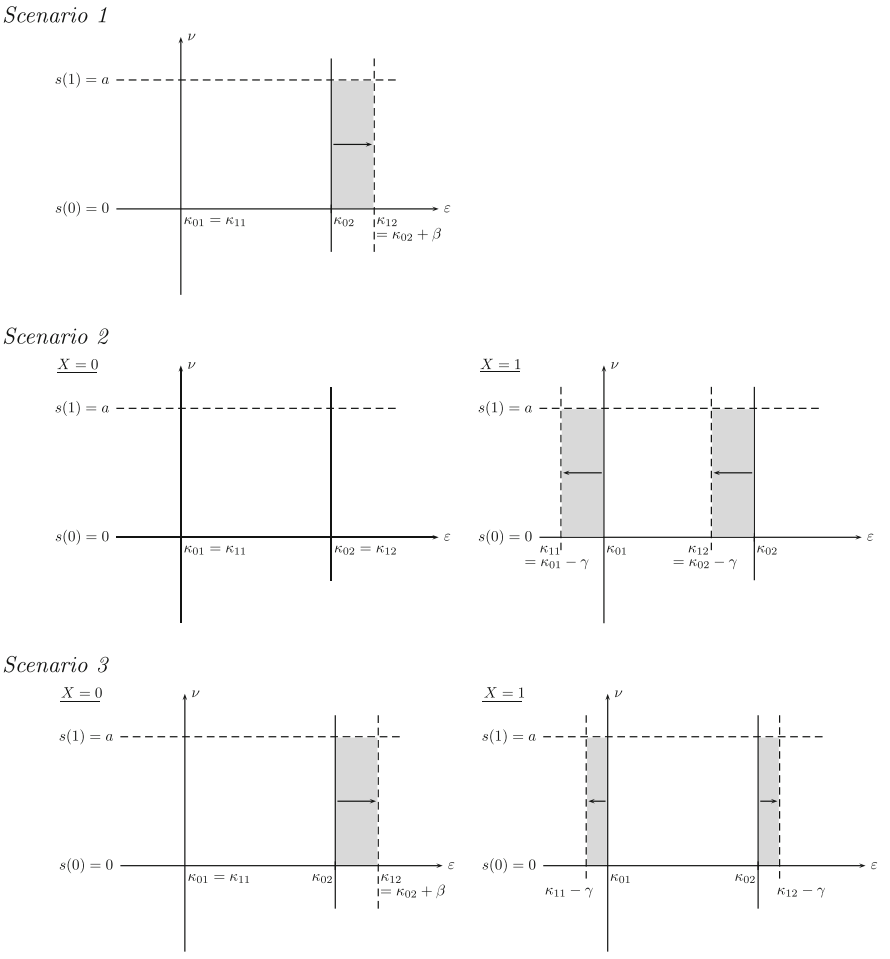

Fig. 1 Graphical illustration of model setup

threshold parameters are not affected by the treatment, but the covariate effect differs for the two potential outcomes. Since X is binary, this corresponds to a symmetric shift in the thresholds if $X = 1$, and no treatment effect if $X = 0$. Scenario 3 assumes that $\beta = 0.2$ and γ is varied from -0.5 to 0.5, and thus asymmetric effects on the threshold parameters are combined with different covariate effects. In each scenario, the strength of the instrument Z is varied with α ranging from 0 to 6.

Figure 1 shows the setup for each scenario. The diagrams draw the (ε, ν)-space with threshold parameters for the treatment selection and the potential outcomes as indicated. In Scenario 1, the treatment does not affect the probability of observing outcome 1. If $\beta > 0$, then the probability of outcome 2 increases with the treatment, the probability of outcome 3 decreases. If $\beta < 0$ then the effects on the probabilities of outcomes 2 and 3 are the opposite. If $\beta = 0$, then the treatment does not affect the outcomes.

The second panel of Fig. 1 draws the threshold parameters as generated by scenario 2 conditional on $X = 0$ (left diagram) and conditional on $X = 1$ (right diagram). The left diagram indicates that the treatment does not affect outcomes conditional on $X = 0$. If $X = 1$, then the effect of the treatment on the thresholds is symmetric. According to the logic of constructing nonparametric bounds, the sign of the treatment effect on the outcome probabilities of the lowest and the highest category can be identified (which is positive or negative depending on the sign of the shift), but nothing (except for the zero effect) can be said about the effect on the middle category, without imposing further assumptions.

In scenario 3, the treatment effect conditional on $X = 0$ is analogous to scenario 1. Conditional on $X = 1$ the treatment effect on the probability of the lowest outcome is determined by the sign of β. For the highest category the effect will depend on the net shift of β as opposed to γ. In the simulation I will keep β fixed at 0.3 such that the cut-off value is $\gamma = 0.3$. The sign of the effect on the middle category can be identified if the upper threshold is shifted to the right and the lower threshold is shifted to the left. With β fixed at 0.3 this case is obtained if γ is in the interval $[0, 0.3]$.

Figures 2, 3, and 4 summarize the results for the bounds on the average treatment effects. In each case, I report the Heckman and Vytlacil (2001a,b) bounds (henceforth HV bounds) under monotone treatment selection, and the bounds derived in Proposition 1 and 2. Consider Fig. 2. The diagrams correspond to the three categories with the results for category 1 shown in the top diagrams, the results for category 2 shown in the middle diagrams, and the results for category 3 shown in the bottom diagrams. In the diagrams on the left side, the horizontal axis represents the values of β with α fixed at 2. In the diagrams on the right side, the horizontal axis represents the values of α with β fixed at 0.2. The true effect in each case is indicated by the thick black line. The small grey circles and triangles correspond to the HV bounds, the large hollow circles and triangles to the bounds of this paper.

As expected, the treatment effect is point-idenitified to be zero by Proposition 1 for the lowest category, irrespective of β, and is point-identified to be zero if $\beta = 0$ for the other categories. The width of the bounds is substantially smaller than the width of the HV bounds. Furthermore, the bounds always identify the sign of the treatment effect. With increasing strength of the instruments the bounds get tighter (see the diagrams on the right side), but point identification is not achieved (even in the limit). The reason for that is the discrete data setup of the example. Even if the treatment status can be perfectly predicted for $Z = 1$, this does not hold conditional on $Z = 0$, and therefore the correlation between Z and D has an upper bound. If Z and D are perfectly correlated, then point identification can be achieved.

The bounds in scenario 2 (Fig. 3) reveal similar results as in scenario 1 for the variation in γ. I only show the bounds for $X = 1$, the case $X = 0$ is trivial. The variation in γ implies a symmetric shift in the thresholds, and the nonparametric bounds therefore do not yield tighter bounds in the middle category, except for the zero treatment effect, than the HV bounds. Again, strengthen the instruments yields tighter bounds, although point-identification is prohibited by the model structure.

The interesting part of scenario 3 (Fig. 4) is the construction of bounds in the second category (the case $X = 1$ is shown, the case $X = 0$ is analogous to scenario 1). Recall that β is fixed at 0.3 and only γ is varied. While the thresholds are shifted in

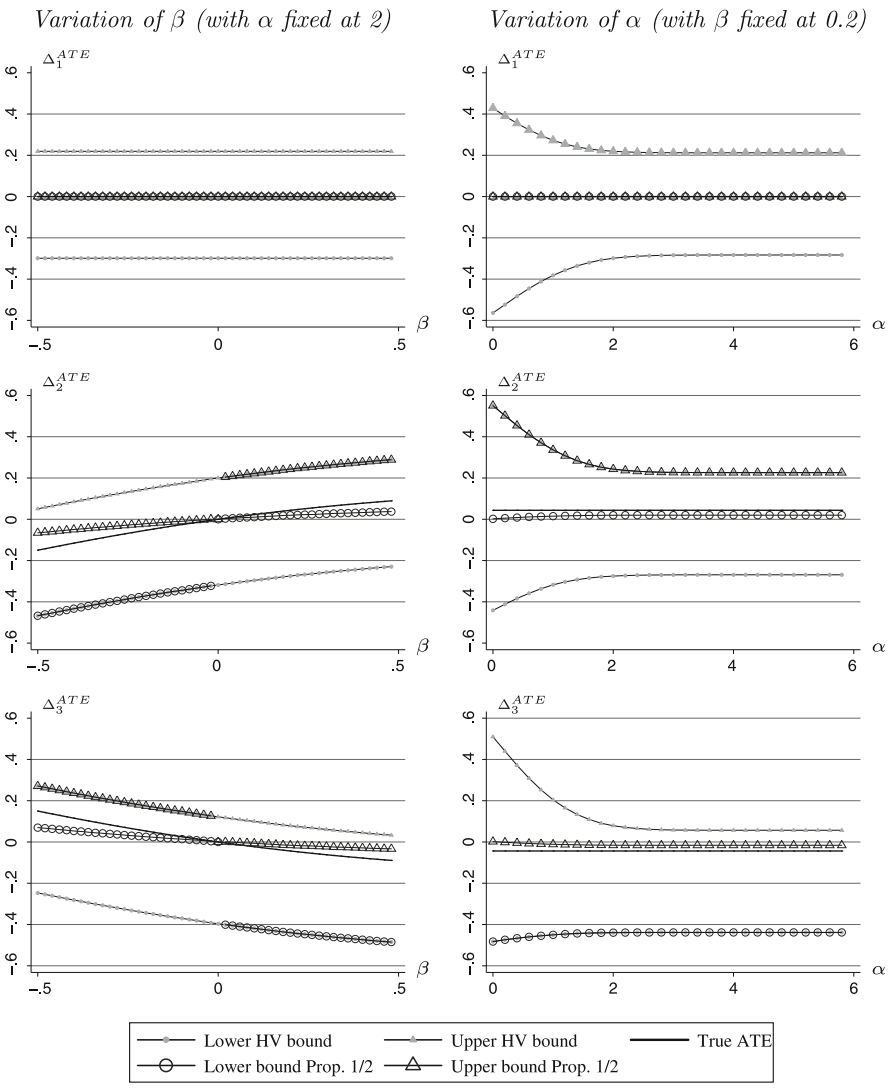

Fig. 2 Bounds on the average treatment effect by category—scenario 1

the same directions for $\gamma < 0$ and for $\gamma > 0.3$, the shift is asymmetric in the interval $[0, 0.3]$. Therefore the bounds derived in this paper are tighter than the HV bounds, and also identify the positive sign of the treatment effect. It is interesting to note from this example that the treatment effect can be zero (for $\gamma \approx -0.248$), but the shift in the upper and lower thresholds goes in the same direction and the nonparametric bounding strategy therefore cannot separate the effects caused by the threshold shift from the effects caused by the shape of the underlying distribution.

As a final aspect, the simulation captures the important practical issue of varying (or summarizing) the number of categories of the outcome variable. Consider again

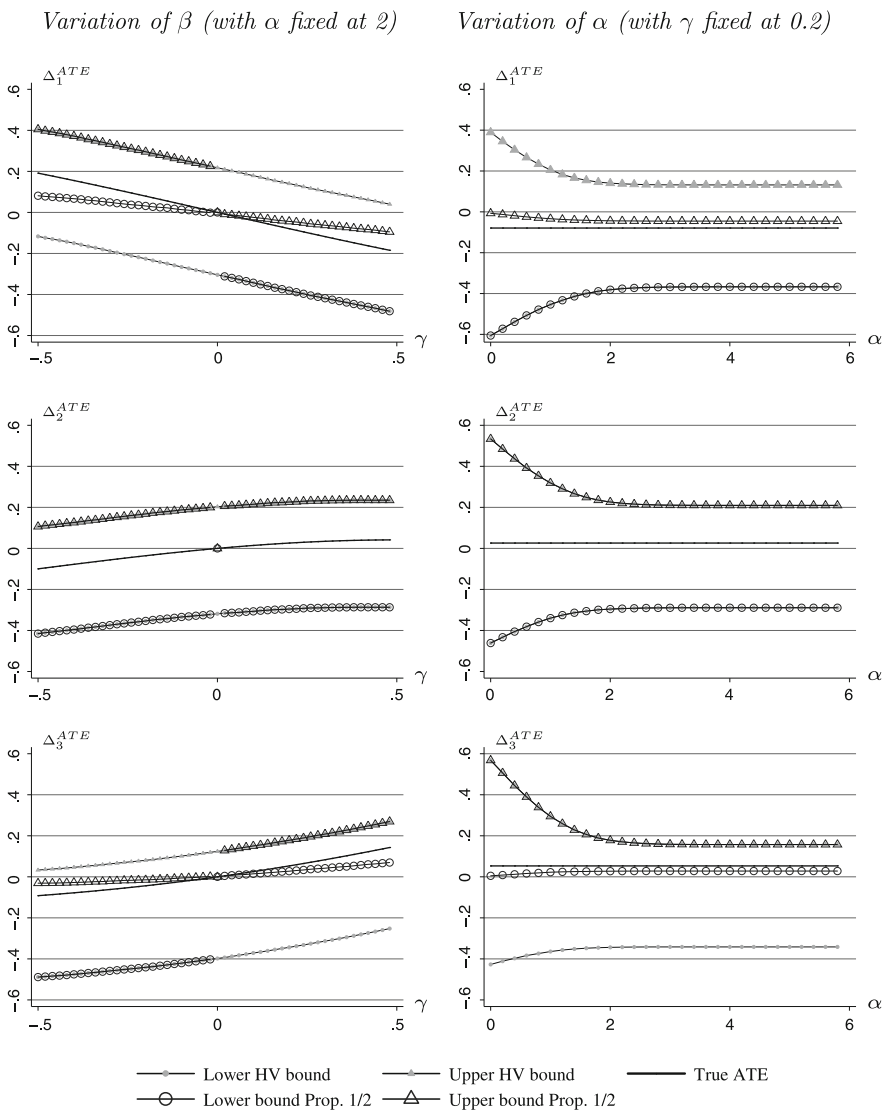

Fig. 3 Bounds on the average treatment effect by category—scenario 2

scenario 1 above with three outcomes (two threshold parameters) and an asymmetric treatment effect on the thresholds. Suppose the researcher collapses categories 2 and 3 and analyzes the effects of the treatment on the new binary outcome variable. Given the model structure and the assumptions of scenario 1, the treatment does not affect outcome 1, and outcomes 2 and 3 are affected by the same absolute magnitude but with opposite signs. If the latter two outcomes are collapsed, then the treatment effect on the joint category is zero, which is point-identified by Proposition 1 (though not by the HV bounds, see Fig. 5).

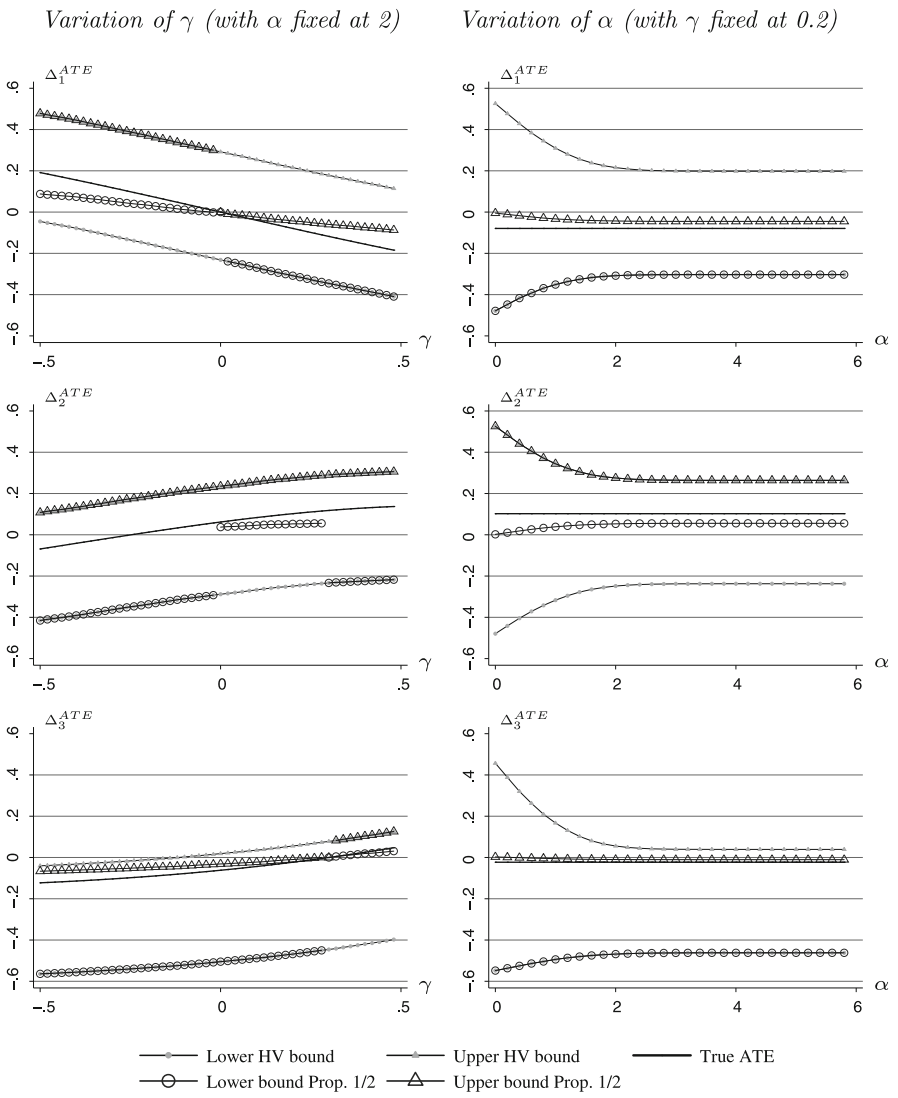

Fig. 4 Bounds on the average treatment effect by category—scenario 3

Now suppose the researcher collapses categories 1 and 2 and analyzes the joint category against category 3 (Fig. 6). While the results for the latter category remain unchanged compared to the case of three outcomes, the HV bounds for the former joint category are shifted upwards. This can be explained by the fact that the probability of outcome 1 in state 1 given $Z = 1$ is larger than the probability of outcome 1 in state 0 given $Z = 0$ under the assumptions of the scenario. The width of the bounds, however, remains unchanged. The bound $P(Y = y|Z = z^u) - P(Y = y|Z = z^l)$ of Proposition 1 (which can be either be an upper or lower bound depending on the

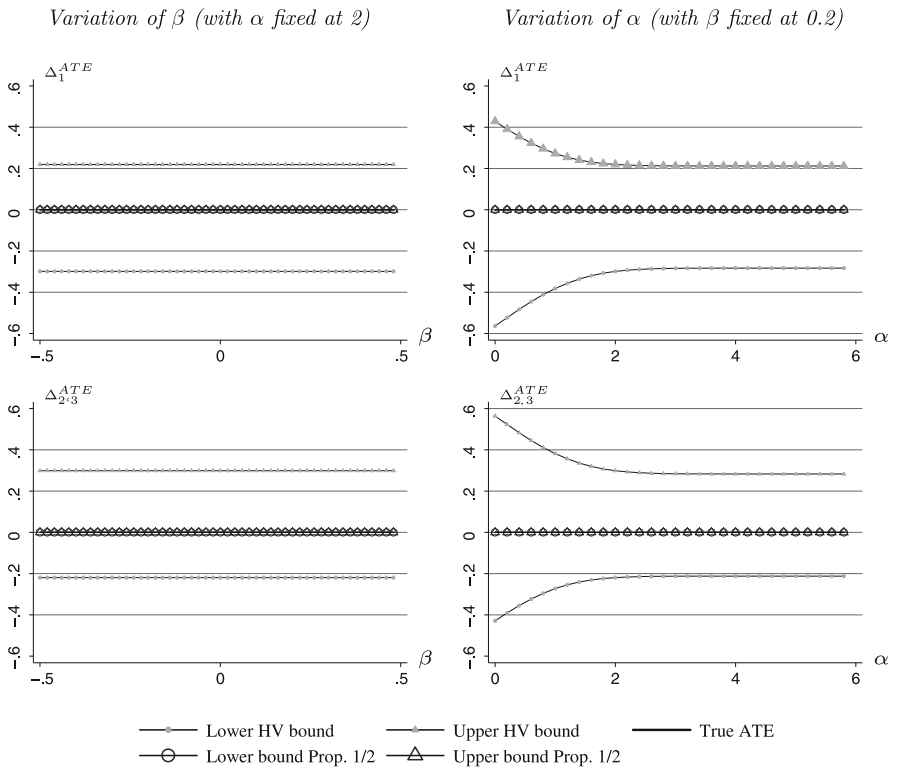

Fig. 5 Bounds on the average treatment effect—scenario 1, joint categories 2 and 3

magnitude of β) is not affected by the collapse of categories 1 and 2 because the treatment does not affect category 1. Thus, depending on the magnitude of β Proposition 1 may either yield tighter or wider bounds if the categories 1 and 2 are summarized in a single category.

To sum up, while the results of the paper still hold if subsequent categories are summarized into a single category, such a procedure does also summarize the effects of the treatment on these categories. This may or may not yield tighter bounds on the average treatment effects of the remaining categories. If the ultimate interest is in the effect of the treatment on the entire outcome distribution, then collapsing categories creates a loss of information.

7 Conclusion

The properties of ordinally measured variables, in a strict sense, require a shift in focus from mean treatment effects to distributional treatment effects. Parametric ordered response models to estimate such effects already exist and are typically based on threshold crossing mechanisms. This is the first paper, to the best of my knowledge,

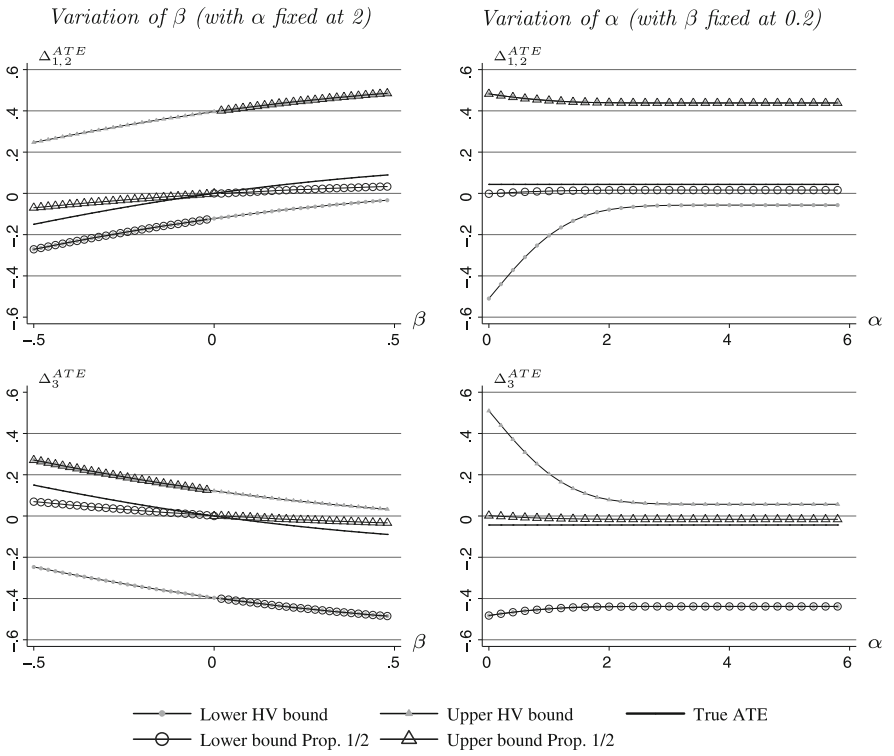

Fig. 6 Bounds on the average treatment effect—scenario 1, joint categories 1 and 2

that discovers the informational content of a threshold crossing mechanism in a nonparametric bounding analysis with ordinal potential outcomes.

The approach taken here complements Shaikh and Vytlacil (2005), who consider the binary/binary case. The extension to ordinal outcomes requires a different identification and bounding strategy, where multiple thresholds need to be taken into account. As a central result, the imposed bounds are never larger than the bounds without imposing the threshold model, and generally smaller. The bounds may identify the sign of the treatment effect, although point-identification except for the zero case generally fails. It is interesting to note that an additional set of parameters becomes available with ordinal outcomes that might be of interest in evaluating the effect of a treatment.

It is left for future research how to conduct inference on the bounds derived here. Special attention to inference is necessary because the usual approach of estimating probabilities by relative frequencies (or replacing population features by sample counterparts) will be inconsistent at the jump points. In particular, the bivariate decision in the first step must be accounted for when constructing a confidence interval with a pre-defined coverage probability.

Acknowledgments I would like to thank Gary Chamberlain, Bo Honoré, Guido Imbens, Michael Lechner, Rainer Winkelmann, Tiemen Woutersen, Kevin Staub, two anonymous referees, and participants at

meetings in Boston, Cambridge, Lausanne, Pittsburgh, Warwick, and Zurich for providing valuable comments.

References

Angrist JD, Imbens GW, Rubin D (1996) Identification of causal effects using instrumental variables. J Am Stat Assoc 91:444–455

Bellemare C, Melenberg B, van Soest A (2002) Semi-parametric models for satisfaction with income. Por Econ J 1:181–203

Björklund A., Moffitt R (1987) The estimation of wage gains and welfare gains in self-selection. Rev Econ Stat 69:42–49

Boes S, Winkelmann R (2006) Ordered response models. Allg Stat Arch 90:165–180

Coppejans M (2007) On efficient estimation of the ordered response model. J Econ 137:577–614

Cunha F, Heckman JJ, Navarro S (2007) The identification and economic content of ordered choice models with stochastic thresholds. Int Econ Rev 48:1273–1309

DeVaro J (2006) Teams, autonomy, and the financial performance of firms. Industrial Relat 45:217–269

Greene W, Hensher D (2010) Modeling ordered choices. Cambridge University Press (forthcoming)

Heckman JJ (1997) Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. J Hum Resour 32:441–462

Heckman JJ, Robb R (1985) Alternative methods for estimating the impact of interventions. In: Heckman JJ, Singer B (eds) Longitudinal analysis of labor market data. Cambridge University Press, New York, pp 156–245

Heckman JJ, Robb R (1986) Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In: Wainer J (ed) Drawing inference from self-selected samples. Springer, New York, pp 63–107

Heckman JJ, Vytlacil EJ (1999) Local instrumental variables and latent variable models for identifying and bounding treatment effects. Proc Natl Acad Sci 96:4730–4734

Heckman JJ, Vytlacil EJ (2001a) Instrumental variables, selection models, and tight bounds on the average treatment effect. In: Lechner M, Pfeiffer F (eds) Econometric evaluation of labour market policies. Physica, Heidelberg, pp 1–15

Heckman JJ, Vytlacil EJ (2001b) Local instrumental variables. In: Hsiao C, Morimune K, Powell J (eds) Nonlinear statistical inference: essays in honor of Takeshi Amemiya. Cambridge University Press, Cambridge, pp 1–46

Heckman JJ, Vytlacil EJ (2005) Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73:669–738

Heckman JJ, LaLonde RJ, Smith JA (1999) The economics and econometrics of active labor market programs. In: Ashenfelter O, Card D (eds) Handbook of labor economics, vol 3. North Holland, Amsterdam, pp 1865–2097

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. Econometrica 62:467–476

Klein RW, Sherman RP (2002) Shift restrictions and semiparametric estimation in ordered response models. Econometrica 70:663–692

Lewbel A (1997) Semiparametric estimation of location and other discrete choice moments. Econ Theory 13:32–51

Lewbel A (2003) Ordered responce threshold estimation. Unpublished working paper, Boston College

Li M, Tobias JL (2008) Bayesian analysis of treatment effects in an ordered potential outcomes model. In: Millimet D, Smith JA, Vytlacil EJ (eds) Advances in econometrics, vol 21: estimating and evaluating treatment effects in econometrics. Elsevier, Oxford, pp 57–91

Luechinger S, Stutzer A, Winkelmann R (2010) Self-selection models for public and private sector job satisfaction. Res Labor Econ (forthcoming)

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60

Manski CF (1990) Nonparametric bounds on treatment effects. Am Econ Rev 80:319–323

Manski CF (1994) The selection problem. In: Sims C (ed) Advances in econometrics: Sixth World Congress, vol 1. Cambridge University Press, Cambridge, pp 143–170

Manski CF (1995) Identification problems in the social sciences. Harvard University Press, Cambridge

Manski CF (2000) Identification problems and decisions under ambiguity: empirical analysis of treatment response and normative analysis of treatment choice. J Econ 95:415–442
Manski CF (2003) Partial identification of probability distributions. Springer, New York
Manski CF (2007) Identification for prediction and decision. Harvard University Press, Cambridge
Manski CF, Pepper J (2000) Monotone instrumental variables: with an application to the returns to schooling. Econometrica 68:997–1010
McCullagh P (1980) Regression models for ordinal data. J R Stat Soc Ser B (Methodol) 42:109–142
McKelvey RD, Zavoina W (1975) A statistical model for the analysis of ordinal level dependent variables. J Math Sociol 4:103–120
Munkin MK, Trivedi PK (2008) Bayesian analysis of the ordered probit model with endogenous selection. J Econ 143:334–348
Neyman J (1923) On the application of probability theory to agricultural experiments. Essay on principles, Sect 9. Stat Sci 5:465–480 (translated)
Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 66:688–701
Scharfstein DO, Manski CF, Anthony JC (2004) On the construction of bounds in prospective studies with missing ordinal outcomes: application to the good behavior game trial. Biometrics 60:154–164
Shaikh AM, Vytlacil EJ (2005) Threshold crossing models and bounds on treatment effects: a nonparametric analysis. NBER Technical Working Papers t0307
Stewart MB (2004) Semi-nonparametric estimation of extended ordered probit models. Stata J 4:27–39
Vytlacil EJ (2002) Independence, monotonicity, and latent index models: an equivalence result. Econometrica 70:331–341
Vytlacil EJ (2006) A note on additive separability and latent index models of binary choice: representation results. Oxf Bull Econ Stat 68:515–518